Title: Evaluating the Impact of Enrollment Prompts on User Engagement: A Data-Driven A/B Test for Udacity Course Trials

Goal:
To reduce the number of students leaving the free trial due to time constraints while maintaining the overall completion rate.

Experiment Design

Metric Choice

Invariant metrics	Evaluation metrics
Number of cookies	Gross conversion
Number of clicks	Retention
	Net Conversion
	Number of user-ids
	Click-through-probability

Measuring Standard Deviation

The standard deviation formula is given as $\sigma = \sqrt{(p(1-p)/n)}$. The calculations were carried out in this spreadsheet. This calculation was carried out analytically.

9		probability	**Standard dev**
10	Gross conversion	0.2063	**0.0072**
11	Retention	0.5300	**0.0194**
12	Net Conversion	0.1093	**0.0055**
13			

—----------

Question:
Do you expect the analytic estimates to be accurate? That is, for which metrics, if any, would you want to collect an empirical estimate of the variability if you had time?

Response:
Gross Conversion: The probability of enrollment given a click is relatively high (20.625%), and the sample size (3,200 clicks per day) is large. For gross conversion, the analytic estimate should be fairly accurate because large sample sizes tend to produce reliable standard deviations when assuming a binomial distribution.

Retention may benefit from empirical validation given its smaller sample size.

Net Conversion, with a lower success probability, is more prone to variability, so collecting empirical estimates would provide a more accurate understanding of this metric's variability.
—----------

Sizing

Number of Samples vs. Power

Using the formula below, the respective power for each evaluation

$$n = \frac{2(\sigma^2)(Z_{\alpha/2} + Z_{\beta})^2}{dmin^2}$$

Where:

- n is the sample size per group.
- σ is the standard deviation of the metric.
- $Z_{\alpha/2}$ is the z-value for the significance level (α = 0.05, so $Z_{\alpha/2}$=1.96).
- Z_{β} is the z-value for the power level (β = 0.2, so Z_{β}=0.84).
- dmin is the minimum detectable effect size.

The results are given below.

	probability	Standard dev	pageview power per group	Total pageviews for both groups
Gross conversion	0.2063	0.0072	8	**16**
Retention	0.5300	0.0194	59	**118**
Net Conversion	0.1093	0.0055	8	**17**
Total				151

To adequately power the experiment, we need to ensure we have enough pageviews for each metric. A total of 151 pageviews are needed.

Duration

To calculate the % of traffic to divert to the experiment, we need to focus on those who click "Start free trial" (3,200 cookies)" The total number of pageviews we need for both groups is **151**.

1	Unique cookies to view course overview page per day:	40000
2	Unique cookies to click "Start free trial" per day:	3200
3	Enrollments per day:	660
4	Click-through-probability on "Start free trial":	0.08
5	Probability of enrolling, given click:	0.20625
6	Probability of payment, given enroll:	0.53
7	Probability of payment, given click	0.1093125

% of traffic to divert = no of pageviews required for both groups/ unique cookies who click "Start free trial"

= 4.73%

4		
5	% of traffic diversion	4.73%
6		

Risk consideration

The change in the experiment doesn't seem too risky (it's a change to messaging about time commitment), so there's no strong reason to avoid diverting all traffic.

Exposure

With 4.7% of traffic diverted to the experiment, it would take **less than a day** to collect enough data for the sample size based on the analytic estimates of variance.

Thus, the experiment would run **very quickly**, and there is no need to reconsider any earlier decisions regarding the sample size or traffic diversion percentage.

Experiment Analysis

Sanity Checks

This spreadsheet shows the cleaned data needed to compute the above metrics, broken down day by day. This includes both the control and experimental groups.

The descriptions of the columns are provided below:

- **Pageviews**: Number of unique cookies to view the course overview page that day.
- **Clicks**: Number of unique cookies to click the course overview page that day.
- **Enrollments**: Number of user-ids to enroll in the free trial that day.
- **Payments**: Number of user-ids who who enrolled on that day to remain enrolled for 14 days and thus make a payment. (Note that the date for this column is the start date, that is, the date of enrollment, rather than the date of the payment. The payment happened 14 days later. Because of this, the enrollments and payments are tracked for 14 fewer days than the other columns.)

Conducting sanity checks for the invariant metric (page views) at a 95% confidence interval for the control and experiment groups. The result below can be found in this Google Sheet above or you can also check here.

Sanity Check on Pageviews (Invariant metric):

Note that all these calculations was carried out in the spreadsheet. I included the calculations here to make it easy to follow.

	Invariant metric							
Metric	Total pageviews	Average pagevie	probability (randoı	Standard error	Margin of error	CI- lower	CI- upper	Sanity check
Pageviews	690203	345101.5	0.5	0.0006	0.0012	344273.2564	345929.7436	Pass

First, I conducted sanity check on the pageviews to verify if randomization between the control and experiment groups was successful. I used a binomial test for this.

Steps for Sanity Check:

1. **Null Hypothesis**: The pageviews should be split equally (50/50) between the control and experiment groups.
2. **Calculate Total Pageviews**:
    - Control Group Total Pageviews: 345,543
    - Experiment Group Total Pageviews: 344,660
    - Combined Total Pageviews: 345,543+344,660= 690,203
3. **Expected Pageviews per Group**: Since the split should be random, the expected pageviews for each group are:

$$\text{Expected Pageviews per Group} = 690{,}203 / 2 = 345{,}101.5$$

4. Standard Error (SE): The standard error is based on the binomial distribution:

$$= \sqrt{p * (1-p)} / \text{ Combined Total Pageviews}$$

Since the cookies are randomly selected, p =0.5.

$$\text{Thus, SE} = 0.0006$$

5. **Margin of Error (ME)**: For a 95% confidence level, the margin of error is:

$$ME = 1.96 \times SE$$
$$= 1.96 \times 0.0006 = 0.0012$$

6. Confidence Interval for Pageviews:

$$CI = 345{,}101.5 \pm (0.0012 \times 690{,}203) = 345{,}101.5 \pm 828.24$$

So, the confidence interval is approximately: [344,273 to 345,929]

7. Observed Values:

- **Control Group Pageviews**: 345,543
- **Experiment Group Pageviews**: 344,660

The control group's pageviews (345,543) fall within the confidence interval [344,273 to 345,929], so the sanity check **passes**.

However, this is not sufficient enough to launch the features. Further analysis needs to be conducted on the evaluation metrics to guide our actions.

Check for Practical and Statistical Significance (Evaluation metrics)

Evaluation metrics used:

Gross Conversion;
Retention; and
Net Conversion.

These three metrics align with the hypothesis and goals of the experiment. They measure user interest their level of commitment and the likelihood of succeeding in the course.
The calculations that we'll be using have been carried out in the spreadsheet here.

Definition:
In this case, a metric is **statistically significant** if the confidence interval does not include 0 (that is, one can be confident there was a change), and it is **practically significant** if the confidence interval does not include the practical significance boundary (that is, one can be confident there is a change that matters to the business).

The spreadsheet below shows the calculated confidence interval (CI) for the three evaluation metrics.

Metrics	Metric differences	P pool	Standard Error	Margin of error	CI- lower	CI- upper	d-min	Statistically Significant?	Pratically Significant?
Gross conversior	-0.0125	0.1271	0.0028	0.0055	-0.0180	-0.0070	0.0100	Yes	Yes
Net Conversion	-0.0030	0.0702	0.0021	0.0042	-0.0072	0.0012	0.0075	No	Yes
Retention	-1.2936	0.5519	0.0117	0.0230	-1.3166	-1.2706	0.0010	Yes	Yes

Sign test

Sign test	No of trials	No of success	Two-tail P value	Two-tail P value< α (0.05)
Gross conversion	23	4	0.0026	Yes
Net conversion	23	10	0.6776	No
Retention	23	13	0.6776	No

Recommendations

Metrics Overview:

- **Gross Conversion:**
  - Statistically significant (both in the original test and the sign test).
  - Practically significant.

- **Net Conversion:**
    - Statistically **not significant**, but practically significant.
    - Sign test also confirms no statistical significance (p-value > α).
- **Retention:**
    - Statistically and practically significant in the original analysis.
    - Sign test indicates **no statistical significance** (p-value > α).

Recommended Action:

Digging Deeper and Running a Follow-Up Experiment. While gross conversion is a strong positive signal, the uncertainty around net conversion and retention suggests that further investigation and refinement of the experiment are needed before a full launch.

Digging Deeper:

- **Focus on retention and net conversion**: Both are practically significant but lack strong statistical significance. Investigating why retention and net conversion didn't reach statistical significance could reveal insights about user behavior.
- Different user segments can be analyzed (e.g., users who enrolled vs. those who accessed free materials) or time periods within the experiment to see if certain groups were more impacted.

Running a Follow-Up Experiment:

- Given that **net conversion** didn’t reach statistical significance but is practically significant, running a follow-up experiment could help to gather more data. Increasing the sample size or extending the experiment duration could provide clearer results on whether net conversion will improve and whether retention shows more definitive significance.
- This follow-up could also refine messaging or UX elements to improve user decisions at key stages like checkout and post-trial conversion.